An Extensive Survey on 3D Face Reconstruction Based on Passive Method

Manasa Sandeep[1], C. Nandini[2]

[1]Research Scholar, Assistant Professor, Department of Computer Science and Engineering, Dayananda Sagar Academy of Technology and Management, Bengaluru, Visvesvaraya Technological University, Belgaum, Karnataka, India- 590018.
[2]Professor and Head, Department of Computer Science and Engineering, Dayananda Sagar Academy of Technology and Management, Bengaluru, Karnataka, India-560082

---***---

Abstract - *3D Reconstruction of a face image is gaining more attention from the last few years due to its vast applications. 3D reconstruction of an image provides a realistic image for analysis. Facial Image Reconstruction of a 2D image gives details on face geometry, texture, depth, color etc. which are used in diverse industries ranging from simple gaming and animation applications to complex medical imaging and biometrics applications. This work attempts to compare the work done in recent years based on passive methods. This paper concentrates majorly on the models that use 3DMM, SFS, SfM, deep learning approaches or the combination of these methods. The paper also discusses some of the available 3D face databases.*

Key Words: 3D face Reconstruction, Passive Methods, 3DMM, Deep Learning

1.INTRODUCTION

Face is a very important part of the human body that provides an individual identity. Human beings can perceive any object in their surroundings as a 3D model due to the presence of a complex optic nerve system. Computer Vision and Machine Learning field has advanced so much in technology that we can represent an image in 3D in computers. But facial Image analysis faces many challenges as the human face keeps changing with expressions, emotions, age, facial hair and other external factors. In addition to these challenges, 3D face reconstruction should also deal with, pose, illumination and occlusion related problems. This gives a wide scope for the researchers to work in this field. The reconstructed 3D faces find its applications in a broad range of applications including forensics, Virtual reality, animation, movies, medical imaging, telecommunication, biometrics, Human Computer Interaction etc. Table I details the different application areas where 3D reconstructed faces are used.

1.1 Applications

Field	Possible Applications	Advantages
Forensics	➢ 3D Modeling of suspects or culprits faces from ○ 2D images ○ Sketches available ○ Images obtained from CCTV footage ➢ Faces reconstructed from cranial remains of dead bodies	Provides individual identification of a person and helps in solving Crime
Biometrics	To provide Access control for ➢ Homes and Offices ➢ Mobiles and Laptops	Provides Security
Medical	Used in Plastic Surgery planning ○ Reconstructive Surgery after burn or an accident ○ Cosmetic Surgery to change facial features	Helps doctors and Patients in providing better understanding of the surgery outcome
Entertainment	➢ Interactive Virtual Reality Systems ➢ Animation in Gaming and Movies	Identifies human expression, and actions

Table 1: 3D Reconstructed Faces Application Areas

2. 3D FACE RECONSTRUCTION

Existing Methods

Active Reconstruction Method and Passive Reconstruction Method are the two methods available for 3D Face Reconstruction.

2.1 Active Method

Active method involves projection of electromagnetic waves like X rays, laser, visible light, ultrasound or microwaves on the object to be reconstructed. The projected waves are reflected back by these objects. The reflected wave reaches back to the source and the 3D model is constructed based on phase change and time taken by the reflected wave to reach the source back. External factors such as scattering, noise, absorption by the medium etc. may affect the constructed model. Hence it requires a controlled environment. Cyberware 3D laser Scanner is one such example. RGB- Depth cameras like Kinect are easier to use and cheaper but lack in quality of the scan.

2.2 Passive Method

Though Active methods give high quality scans, they need expensive machinery and the objects are exposed to radiation. Passive methods construct the 3D model without obstructing directly with the target objects. It uses 2D images captured from digital cameras to estimate the face geometry. The algorithms used will single image or sequence of images as input.

3. PASSIVE METHOD BASED 3 DIMENSIONAL FACE RECONSTRUCTION

3.1 General Steps in Facial Reconstruction

Figure 1 represents the general steps of the 3D face reconstruction and the next section explains each of the steps in detail

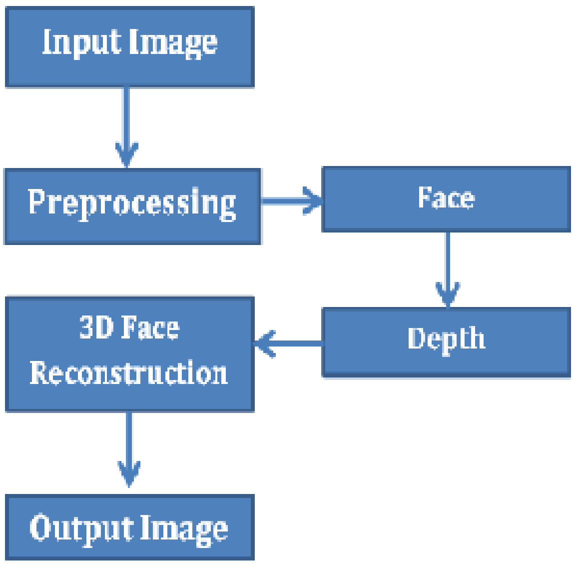

Fig- 1: General steps of the 3D face

3.1.1. Input Image

2D Image is taken as input. Input Image can Single 2D Image or more than one image like frontal and side view etc. or a sequence of Images taken from video like CCTV etc. based on the approach and the algorithm used.

3.1.2 Preprocessing

Once the input image is obtained, it is always preprocessed. Presence of noise, shadows, Occlusion, poor lighting may affect accuracy of the reconstructed image. This step removes any noise present in the input image, performs contrast and intensity adjustment, repairs if the input image is corrupted from shadow, poor lighting or occlusion.

3.1.3 Face Localization

The process of identifying the face in the given picture is called face localization. This involves removal of background and identifying the facial parts like eyes, mouth, nose using different face alignment algorithms.

3.1.4 Depth Estimation

Depth finding plays a major role in 3D reconstruction as the depth feature gives the realistic look to the reconstructed image. Depth estimation can be performed by building the depth map.

3.1.4 3D Face Reconstruction

Once the face is identified and localized in the given input image, next step is to reconstruct the face by using 3D face reconstruction approach. Several such approaches exist.

3.2 Challenges Involved

Several challenges are involved in reconstructing an image from 2D input image. The challenges are due to many reasons. The following are few reasons that make 3D face reconstruction a difficult task.

- Change in Facial Expression of an individual
- Problems arises due to change in pose of the neck, face and head
- Lighting Problem, Noise and other environmental factors that affects the input image
- Occlusion Problem - when part of a face is covered from some obstacles

3.3 General Approaches

There are different approaches available for 3D reconstruction of face. By referring to the literature, this paper categorizes them into four types namely:

- The 3 D Morphable Model Approach which is popularly known as 3DMM
- Reconstruction from Shading
- Reconstruction from Video
- Deep Learning Methods known as DL Methods

4. THE 3 D MORPHABLE MODEL APPROACH

This method uses a generic 3D face model. It fits the given input image to that generic model and reconstructs the face. This is a passive method that takes one or more images as input. The face model used could be 3DMM or any other generic model. Few other facial models are Basel Face Model (BFM) [1], Surrey Face Model (SFM) [2] etc. Blanz and Vetter proposed 3D Morphable Model which is commonly known as 3DMM is the widely used model [3]. The idea behind 3DMM is that the linear sum of basic functions can be used to represent the 3D facial shape. It allows reconstructing the shape and albedo by solving a nonlinear optimization problem. This approach gives the best results when the input image is similar to that of the general image used.

5. RECONSTRUCTION FROM SHADING

Basically, Reconstruction from Shading is the process of using one input image of the object to compute the 3D shape of that object. 3D face reconstruction using the SFS approach uses intensity variation of the image input to reconstruct the 3D face. This concept was initially given by B. K. P. Horn in 1989 [4]. This method is based on the Lambertian model. SFS obtains good 3D information from smooth surfaces. When used with face images, it works very well in estimating the depth feature, but fails in handling the areas where the texture changes are not smooth like nose, eyes etc. Researchers handle this in different ways like combining this approach with other approaches. This improves the accuracy of the resulting model etc.

6. RECONSTRUCTION FROM VIDEO

In this method, continuous frames from a single source like video input is considered as the input. 3D reconstruction of face is done by using this input. The images obtained by the input video sequence are analyzed to estimate the 3D face output. The correspondence between the available different input images is used to find the output Image. Features from the face are selected and they are tracked from one image to the next, which helps in finding the correspondence that exists between them and constructs the 3D face.

7. DEEP LEARNING METHODS

This method follows the Machine learning approach and the solution is normally developed in two phases, phase one is the learning phase and the second being the reconstruction phase. The model is rigorously trained in the training phase. The prior information known is used to train the model. Prior knowledge is encoded in the form of weights in the trained network. Using this training, the system learns to find the mapping. It learns to map between the 2D input image and the 3D face. The success of this method depends on the number of training data set used, type of the network used, number of layers used in training phase etc. Table 2 compares the work done on 3D face Reconstruction in recent years based on passive methods

Sl. No	**Reference Number**	**Year**	**Approach Used**	**No. of Input Images Considered**	**Input Image Environment (Wild/Controlled)**	**Methodology Used**	**Research Gap**	**Recognition Accuracy**
1	[5]	2003	3DMM	Single	Controlled Environment	Computer Graphics simulation is combined with Deformable 3D model. Projection and illumination are used. Shape and texture are used as intrinsic parameters. Both Intrinsic and extrinsic parameters are independent of each other.	Highly dependent on the database used	Hit Ratio CMU-PIE database: 77.5% FERET database: 87.9% False Alarm Rsate:1%
2	[6]	2017	Reconstruction from Video	Single	wild	Structure from motion is used which is named as “Prior constrained structure from motion (PCSfM)”. It	Concentrates on sequence of frames/ video input.	3D Root Mean Square Error: Without considering pose and Expression:

						uses the 3DMM approach to find the prior knowledge of shapes. Video input is used to find the landmarks of the input face and fitting the prior knowledge to estimate the face shapes		2.09 By considering pose and Expression: 1.92
3	[7]	2018	3DMM + Shading	Single Image	wild	The approach uses different steps, first step gives a 3D face which is smooth. This is generated using "Example-based bilinear face model". Then medium face shape is obtained by refining bilinear face model using photometric consistency constraints. Finally, fine geometric details are recovered from medium face by using shape-from-shading method.	Gives good result only when coarse face model is similar to the ground-truth overall shape	3D Root Mean Square Error: Expression Happy: 1:71 ± 0:34 Surprise : 2:05 ± 0:49 Disgust: 1:98 ± 0:42
4	[8]	2019	Deep Learning	Single	Wild	A MobileNet based fusion network is proposed. This method has encoder that estimates the following parameters of a face model: • Pose • Identity • Expression coefficients • Coefficients of a shape correction model Next, the final output face shape is given by the decoder module.	The proposed algorithm is not tested with the landmarks of different databases and fine detail face reconstruction. It doesn't give satisfactory results when we don't fix the position on the face	Average Normalized mean error (NME) when correction is not used: 3.41 When correction is used: 3.37
5	[9]	2019	3DMM + Deep Learning	Single Image	Wild	3DMM fitting approach is used in combination with the deep features that are identical bytraining a GAN- "Generative Adversarial Network" on a large scale 3D texture dataset	Though successfully reconstructs the face in presence of Self- self-occlusions and facial hair and strong	Shape Estimation Mean and Standard Deviation GANFit++ Cooperative Condition: 0.94 ± 0.17 Indoor Condition: 0.92 ± 0.14

							illuminations, there is a scope for improving absolute error	Outdoor Condition: 0.94 ± 0.19
6	[10]	2019	Deep Learning	Single /Multiple Images	Wild	The proposed architecture reconstructs the 3D face based on siamese neural network using one or multiple images	Gives dedicated architecture. It is not capable of performing efficiently the extraction and merging of the 3D information from multiple input views	Average Root Mean Square Error of the two models proposed MV Add Module: 2.43 MV Concat Module: 2.33
7	[11]	2020	3DMM	Multiple Images	Wild-Gradient Correlation objective function is used	Landmark Features are estimated. Setting up of Interval for basis coefficients. Sparse landmark detectors are used	The efficiency of the "Inequality constrained minimization algorithms" is not evaluated	Normalized mean error for Synthesized dataset for 68 Landmarks Single-frame: 0.036 Multi-frame: 0.032
8	[12]	2020	Shading	Single	Wild	Basel face model is used. Facial features extracted are subjected to scale-invariant feature transform. Multivariate Gaussian distribution is used to calculate depth	Tested only for Wild Home- LFW database and it uses faces that are labelled. Can't handle when an image input is givenrandomly.	RMS error for Image 1: 0.8866 Image 2: 0.8689 Image 3: 1.6468
9	[13]	2020	3DMM + Deep Learning	Multiple Images	Wild	System uses one encoder and two decoder networks. The input 2D face image is encoded by encoder network according to • Identity • Expression • Pose etc. The 3D face shape is	The reconstructed face shape depends on stage 1 output. The stage 1 is called "foundation	Mean error comparison of average over three-view: Cooperative: 1.16±0.29 Indoor: 1.24±0.30 Outdoor: 1.22±0.28

						regressed by the first decoder network and the second regress the albedo. Then the system renders the facial image back and matches the input image.	face shape". This stage uses linear 3DMM which doesn't handles the wild condition when the input contains diversity of poses, occlusions, expressions, lightings etc.	
10	[14]	2020	3DMM + Deep Learning	Single	Wild	3DMM encoder which is coarse is used for regressing the parameters of 3DMM, then face of the input image and the texture of the 3DMM is unwrapped into the UV space. Here the faces are aligned precisely.	Fails to handle the input that includes • large expression • occlusion • large pose.	POINT-TO-PLANE ERROR Indoor Cooperative: 1:35±0:31 Outdoor: 1:25±0:21 RMSE: 1:81±0:43

TABLE -2. Comparison of work done on 3D face Reconstruction

8. 3D FACE DATABASE

Many 3D face databases are available. The availability has increased a lot during last few years. Some of the databases like BFM, FLAME etc. are made public and few others like FaceScape, Florence etc. can be accessed on request for non-commercial purposes. Table III lists a few available 3D face databases.

Sl. No	Database	No of Subjects	Image Source	Age Coverage – years	Ethnicity	Expression	Landmark localized
1	FaceWarehouse [15]	150	Kinect	7 to 80	Wide Range	neutral expression and 19 other expressions	Yes
2	Face Scape [16]	847	multi-view system-68 Cameras	16 to 68	-	20 expressions	Yes
3	BFM2009 [17]	200	Two projectors and three cameras and SLR camera	8 to 62	Europeans	Neutral Expressions	Yes
4	BFM2017 [18]	200	Two projectors and three cameras and SLR camera	16 to 70	European Union	6 expression	Yes
5	FLAME [19]	3800	multicamera active stereo system. It consists of 3 pairs of • Stereo cameras • Color cameras • Speckle projectors • White light LED panels.	wide range	wide range	Includes expressions	Yes

TABLE 3. LIST OF 3D FACE DATABASES

9. CONCLUSION AND DISCUSSIONS

3D reconstruction of the face is a challenging task. 3D face Reconstruction from one input image taken in wild condition is more challenging than using multiple images as input. Most of the work concentrates on two to three expressions (neutral, happy, sad etc.) and two to three conditions (cooperative indoor, outdoor condition etc.) while reconstructing the 3D. Also only few works considers the time complexity and throughput of the proposed methods. The future works can also concentrate on improving throughput, efficiency, time complexity. Through critical discussion on the results of the many models as discussed in this paper, still there is scope to improve the efficiency when the input scenarios contain wide expression, ethnicity and occlusion. There is need for the multimodal approach to improve on various parameters as discussed above to reconstruct the images taken in a more challenging situation to serve for real time applications

REFERENCES

[1] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, T. Vetter, ‘A 3D face model for pose and illumination invariant face recognition’, Proc of IEEE AVSS,pp. 296- 301, 2009 .

[2] P. Huber, G. Hu, J. R. Tena, P. Mortazavian, P. Koppen, W. J. Christmas, M. Ratsch, J. Kittler,’A multiresolution 3D morphable face model and tting framework’, Proc. of VISAPP, pp. 79- 86, 2016.

[3] V. Blanz, T. Vetter, ‘A morphable model for the synthesis of 3D faces’, Proc. of ACM SIGGRAPH, pp- 187-194, 1999.

[4] B. K. P. Horn, ‘Obtaining shape from shading information’ in B. K. P. Horn and M. J. Brooks, MIT Press, Shape from Shading, pages 121–171,1989.

[5] V. Blanz and T. Vetter, ‘Face recognition based on fitting a 3D morphable model’ in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 9, pp. 1063-1074, Sept. 2003.

[6] M. Hernandez, T. Hassner, J. Choi, G. Medioni, ‘Accurate 3D face reconstruction via prior constrained structure from motion’, Computer Graphics. 66 pp. 14- 22, 2017.

[7] L. Jiang, J. Zhang, B. Deng, H. Li, L. Liu, ‘3D face reconstruction with geometry details from a single image’, IEEE Transaction on Image Processing. 27 4756-4770, 2018.

[8] P. Wang, Y. Tian, W. Che, B. Xu, ‘Efficient and accurate face shape reconstruction by fusion of multiple landmark databases’, published in Proc. of ICIP, pp. 335-339, 2019.

[9] B. Gecer, S. Ploumpis, I. Kotsia, S. Zafeiriou, ‘GANFIT: Generative adversarial network fitting for high fidelity

3D face reconstruction', in: Proc. of IEEE CVPR, pp. 1155-1164, 2019.

[10] E. Ramon, J. Escur, X. Giro-i Nieto, 'Multi-view 3D face reconstruction in the wild using siamese networks', published in Proc. of IEEE ICCV Workshops, pp.3096-3100, 2019.

[11] E. Sariyanidi, C. J. Zampella, R. T. Schultz, B. Tunc, 'Inequality-constrained and robust 3D face model fitting', in: Proc. of ECCV, 2020.

[12] H. M. Rehan Afzal, Suhuai Luo, M. Kamran Afzal, Gopal Chaudhary, Manju Khari, Sathish A. P. Kumar, '3D Face Reconstruction From Single 2D Image Using Distinctive Features', IEEE Access 8:180681 – 180689, October 2020

[13] Z. Gao, J. Zhang, Y. Guo, C. Ma, G. Zhai, X. Yang, 'Semisupervised 3D face representation learning from unconstrained photo collections', Proceedings of IEEE pp. 348-349, IEEE CVPR Workshops, 2020.

[14] Y. Chen, F. Wu, Z. Wang, Y. Song, Y. Ling, L. Bao, 'Selfsupervised learning of detailed 3D face reconstruction, IEEE Transactions on Image Processing, 29 pp. 8696-8705, 2020.

[15] Chen Cao, Yanlin Weng, Shun Zhou, Yiying Tong, Kun Zhou, 'FaceWarehouse: A 3D Facial Expression Database for Visual Computing', IEEE Transactions on Visualization and Computer Graphics, Volume: 20, Issue: 3, 2014.

[16] Haotian Yang, Hao Zhu, Yanru Wang, Mingkai Huang, Qiu Shen, Ruigang Yang, Xun Cao, 'FaceScape: a Large-scale High Quality 3D Face Dataset and Detailed Riggable 3D Face Prediction', CVPR 2020.

[17] Pascal Paysan, Reinhard Knothe, Brian Amberg, Sami Romdhani, Thomas Vetter, 'A 3D Face Model for Pose and Illumination Invariant Face Recognition', 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance, 2009

[18] T. Gerig, A. Morel-Forster, C. Blumer, B. Egger, M. Luthi, S. Schoenborn, T. Vetter, 'Morphable face models - An open framework', Proc. of IEEE FG, pp. 75-82, 2018.

[19] T. Li, T. Bolkart, M. J. Black, H. Li, J. Romero, 'Learning a model of facial shape and expression from 4D scans', ACM Transaction. Graphics. 36, 2017.